> "Express Mail" mailing label number:
>
> EV 335896351 US

RANK-BASED ESTIMATE OF RELEVANCE VALUES

Richard Slackman

Field of the Disclosure

[1001] The present disclosure relates to search engines.

Description of the Related Art

[1002] U.S. Patent Application Publication No. 2003/0041054 to Mao et al. discloses a method of merging results lists from multiple search engines. In particular, each of the search engines returns a respective results list based on a search query. A subset of entries in each results list is selected. Each entry in the subset is assigned a scoring value representing how closely the entry matches the search query. Each results list is assigned a representative value based on the scoring values assigned to its subset of entries. A merged list is formed based on each representative value and each scoring value.

BRIEF DESCRIPTION OF THE DRAWINGS

[1003] The present invention is pointed out with particularity in the appended claims. However, certain features are described in the following detailed description in conjunction with the accompanying drawing in which:

[1004] FIG. 1 is a block diagram of a system for merging search results lists;

[1005] FIG. 2 is a flow chart of an embodiment of a method performed by a computer upon receiving a search query;

[1006] FIG. 3 is a block diagram to illustrate an example of acts described with reference to FIG. 2; and

- 1 -

[1007] FIG. 4 is a flow chart of an embodiment of a method of determining a search engine weighting value.

DETAILED DESCRIPTION OF THE DRAWINGS

[1008] Embodiments of the present invention address situations in which some search results have a search-engine-supplied relevance value indicating how closely the result matches a search query, but other search results are absent a search-engine-supplied relevance value. An estimated relevance value is determined for those search results absent the search-engine-supplied relevance value. The estimated relevance value for a search result is based on its rank in a ranked list of the search results, and known relevance values and ranks of at least two others of the search results.

[1009] Further, embodiments of the present invention address meta-search situations in which all search engines queried are not assumed to provide equally relevant results. A search engine weighting factor is determined to weight search results based on their originating search engine.

[1010] FIG. 1 is a block diagram of a system for merging search results lists. The system comprises a computer 10 that receives a search query 12 from a computer 14 of an end user 16. The computer 10 communicates with multiple search engines 20 to perform multiple searches based on the search query 12. For purposes of illustration and example, three search engines 22, 24 and 26 are depicted. Those having ordinary skill will appreciate that an arbitrary number of search engines may be used.

[1011] Communication between the computers 10 and 14, and between the computer 10 and the search engines 20, may be facilitated by a computer network 30. Examples of the computer network 30 include, but are not limited to, the Internet, an intranet, an extranet, a local area network (LAN) and a wide-area network (WAN).

[1012] An embodiment of a method performed by the computer 10 upon receiving the search query 12 is described with reference to FIG. 2. The method can be directed by computer program code stored by a computer-readable medium which causes the computer 10 to perform the acts described herein. Examples of the computer-readable

- 2 -

medium include, but are not limited to, a magnetic medium such as a hard disk, an optical medium such as an optical disk, and an electronic medium such as an electronic memory.

[1013] Reference will be made to FIG. 3 to illustrate an example of the acts described with reference to FIG. 2. Those having ordinary skill should appreciate that the scope of this disclosure is not limited by the example provided in FIG. 3.

[1014] As indicated by block 50, the method comprises submitting the search query 12 to the search engines 20. Each of the search engines 20 performs a search based on the search query 12, and generates a ranked list of search results. Some, and possibly all, of the search results in each list have an associated relevance value determined by its originating one of the search engines 20.

[1015] As indicated by block 52, the method comprises receiving a ranked list of search results from each search engine based on the search query 12. For each list, each search result therein is ranked based on its relevance value to the search query 12. The ranks of the search results in a list may be indicated by listing the search results in descending order of relevance, for example. In another example, the ranks of the search results in a list may be explicitly indicated by a rank value, in which case the ordering of the search results in the list may or may not be relevance-based.

[1016] The act of block 52 may further comprise receiving, from each of the search engines 20, a plurality of actual relevance values for one or more of the search results in its associated list. Each search-engine-supplied, actual relevance value indicates a degree of relevance between the search query 12 and its associated search result. Each of the search engines 20 may have its own scale for relevance values, and/or its own methodology for determining the relevance values. For example, some of the search engines 20 may have a scale from 0 to 100, others may have a scale from 0 to 10, and still others may have a scale from 0 to 1.

[1017] To illustrate an example of the ranked lists, consider the search query 12 comprising "capital gain". FIG. 3 shows three ranked lists 62, 64 and 66 of search results received by the computer 10 from the search engines 22, 24 and 26, respectively, based

- 3 -

on the search query 12. The ranked list 62 includes five search results, each of which having an associated relevance value on a scale from 0 to 100 supplied by the search engine 22. The ranked list 64 includes seven search results, five of which having an associated relevance value on a scale from 0 to 10 supplied by the search engine 24, and two of which having no associated relevance value. The ranked list 66 includes three search results, none of which having an associated relevance value supplied by the search engine 26.

[1018] Referring back to FIG. 2, an act of normalizing the relevance values in the ranked lists of search results is performed, as indicated by block 68. Based on a known scale for each search engine's relevance values, the relevance values are normalized so that each list has the same scale. For example, the relevance values may be normalized so that all have a scale from 0 to 100.

[1019] Continuing with the example of FIG. 3, consider that the relevance values are to be normalized to have a 0-to-100 scale. Since the relevance values in the ranked list 62 are already scaled between 0 and 100, normalized values in block 72 are unchanged from those in the ranked list 62. The relevance values in the ranked list 64, being scaled between 0 and 10, are multiplied by 10 in block 73 to produce normalized values in block 74. Since the ranked list 66 has no relevance values, there is no need to normalize any values to arrive at a list 76.

[1020] Referring back to FIG. 2, the method comprises estimating a relevance value for at least one of the search results absent an actual relevance value, as indicated by block 78. In some embodiments, a relevance value is estimated for each of the search results absent an actual relevance value. The relevance value is estimated based on the rank of its search result in the list, and the ranks and actual relevance values of at least two others of the search results. The relevance value may be estimated using either curve fitting, interpolation or extrapolation.

[1021] In some embodiments, a curve representing relevance as a function of rank is fit to each list having an unknown relevance value. The curve is fit to some or all of the known, actual relevance values in the list. A relevance value for a search result is

- 4 -

estimated by evaluating the fitted curve at the known rank of the search result. If the curve is a line, linear regression can be used to generate a linear relevance vs. rank function. Non-linear relevance vs. rank curves are also contemplated.

[1022] In other embodiments, an interpolation function representing relevance as a function of rank is determined for each list having an unknown relevance value. The interpolation function is based on some or all of the known, actual relevance values in the list. A relevance value for a search result is estimated by evaluating the interpolation function at the known rank of the search result. In one embodiment, an unknown relevance value of a search result is estimated by linearly interpolating between two known relevance values whose ranks bracket (i.e. one rank is greater than, another rank is less than) the rank of the search result. Non-linear interpolation functions are also contemplated.

[1023] Continuing with the example of FIG. 3, since all of the relevance values in the normalized, ranked list 72 are actual, known values, there is no need to estimate any relevance values to arrive at the values in list 82. Two relevance values, 4B and 7B, from the normalized, ranked list 74 are to be estimated 83. In this example, linear least squares regression is used to fit a line to the (rank, relevance) values of (1, 98), (2, 82), (3, 78), (5, 70) and (7, 51). A linear function of relevance $= -7*\text{rank} + 100.8$ results from the linear regression. The estimated relevance value of search result 4B, after rounding, is $-7*4 + 100.8 = 73$. The estimated relevance value of search result 7B, after rounding, is $-7*7 + 100.8 = 51$. A resulting list including both known and estimated relevance values is shown in block 84.

[1024] All relevance values, 1C, 2C and 3C, from the normalized, ranked list 76 are to be determined 85. The computer 10, using its own measure of relevance which may or may not differ from a measure of relevance employed by the search engine 26, determines actual relevance values for at least two of the search results in the list 76. The number of search results for which the computer 10 determines actual relevance values can be based on a desired accuracy. The remaining search results in the list are estimated using curve fitting, interpolation or extrapolation for the actual relevance values determined by the

- 5 -

computer 10. In one embodiment, the computer 10 determines actual relevance values for the most relevant search result (e.g. 1C) and least relevant search result (e.g. 3C) in the list 76. The remaining search results in the list are estimated by linearly interpolating between the most relevant and least relevant search results. For purposes of illustration and example, consider the computer 10 to determine an actual relevance value of 80 for the search result 1C, and an actual relevance value of 55 for the search result 3C. A linear function of relevance = -12.5*rank + 92.5 results from the linear interpolation. The estimated relevance value of search result 2C, after rounding, is -12.5*2 + 92.5 = 68. A resulting list including both actual and estimated relevance values is shown in block 86.

[1025] It is noted that as an option, the computer 10 can determine actual relevance values for all unknown relevance values in a list. However, since the measure of relevance employed by the computer 10 may differ from the measure of relevance employed by the search engine 26, this approach may result in a contradiction between the ranks and the relevance values (i.e. in comparison to one search result, another better-ranked search result may have a lower relevance value). The use of linear interpolation based on the most relevant and least relevant search results in the list mitigates the likelihood of the contradiction.

[1026] Returning to FIG. 2, the method comprises determining an associated weighting value for each of the search engines 20 as indicated by block 88. The weighting value indicates how relevant the search engine or its results are to search query 12. For example, a search engine for a specialized information base that pertains to the search query 12 may be given a greater weighting value than a search engine for a general information base. An embodiment of a method of determining a search engine weighting value is subsequently described with reference to FIG. 4. For purposes of illustration and example, consider the search engine 22 having a weighting value of 8 for the search query 12, the search engine 24 having a weighting value of 5 for the search query 12, and the search engine 26 having a weighting value of 4 for the search query 12.

[1027] As indicated by block 90, the method comprises determining, for each of the ranked lists, an associated weighted relevance value for each of its search results based

- 6 -

on an actual or estimated relevance value of the search result and the weighting value associated with the search engine that provided the ranked list. In one embodiment, the weighted relevance value is a product of the aforementioned actual or estimated relevance value and the aforementioned weighting value.

[1028] Continuing with the example of FIG. 3, the relevance values in the list 82 are multiplied (see block 91) by the weighting value of 8 associated with the search engine 22 to produce a list 92 of weighted relevance values. The relevance values in the list 84 are multiplied (see block 93) by the weighting value of 5 associated with the search engine 24 to produce a list 94 of weighted relevance values. The relevance values in the list 86 are multiplied (see block 95) by the weighting value of 4 associated with the search engine 26 to produce a list 96 of weighted relevance values.

[1029] Returning to FIG. 2, the method comprises combining the search results into a single list, as indicated by block 100. Continuing with the example of FIG. 3, the lists 92, 94 and 96 are combined to form a single list 102.

[1030] As indicated by block 104, the method comprises sorting the single list in descending order based on the weighted relevance values. Continuing with the example of FIG. 3, the single list 102 is sorted 106 based on the weighted relevance values to form an ordered list 110.

[1031] As indicated by block 112, the method comprises presenting at least a portion of the ordered list 110. This act may comprise the computer 10 outputting a signal to be communicated to the computer 14 via the computer network 30. The signal may encode a search results page or pages, such as a Web page or pages or another type of electronic document, that presents the ordered list 110. The signal may include code in a markup language such as Hypertext Markup Language (HTML), Handheld Markup Language (HDML) or Wireless Markup Language (WML) to present the ordered list 110. The computer 14 receives the signal and displays some or all of the ordered list 110 for view by the end user 16. Each of the search results in the ordered list 110 may be presented with a hyperlink to its corresponding electronic document (e.g. a Web page, an image, a

- 7 -

word processing document, or a portable document file). The hyperlink can be selected by the end user 16 to access the corresponding electronic document.

[1032] FIG. 4 is a flow chart of an embodiment of a method of determining a search engine weighting value. As indicated by block 120, the method comprises determining one or more categories associated with the search query 12. Continuing with the example of the search query 12 comprising "capital gain", consider two categories associated with the word "capital". A first category has the following category query terms: capital, payment and financial. A second category has the following category query terms: capital, Washington, capitol and government.

[1033] As indicated by block 122, the method comprises determining an absolute relevance value of each category based on the search query 12 and query terms associated with the category. Continuing with the above example, the search query 12 of "capital gain" is compared to the category query terms of "capital", "payment" and "financial" to yield an absolute relevance value of 66 for the first category. The search query 12 of "capital gain" is compared to the category query terms of "capital", "Washington", "capitol" and "government" to yield an absolute relevance value of 25 for the second category.

[1034] As indicated by block 124, the method comprises determining a relative relevance value for each category by dividing its absolute relevance value by a sum of all absolute relevance values. Continuing with the above example, the relative relevance value for the first category, after rounding in percentage terms, is 66/(66+25) = 73%. The relative relevance value for the second category, after rounding in percentage terms, is 25/(66+25) = 27%.

[1035] As indicated by block 126, the method comprises determining category search engine weighting factors for each of the one or more categories. The category search engine weighting factor indicates how well the search engine is expected to find results in the particular category. Continuing with the example, consider the search engine 22 having a factor of 10 for the first category and a factor of 3 for the second category. Consider the search engine 24 having a factor of 5 for the first category and a factor of 5

- 8 -

for the second category. Consider the search engine 26 having a factor of 2 for the first category and a factor of 9 for the second category. Based on the aforementioned factors, one can view the search engine 22 as being more financially-oriented, the search engine 26 as being more government-oriented, and the search engine 24 as being more generalized.

[1036] As indicated by block 130, the method comprises determining the search engine weighting value based on a sum, over the categories, of each product of its associated category search engine weighting value and its relative relevance value. Continuing with the example, the search engine weighting value for the search engine 22, after rounding, is $10*0.73 + 3*0.27 = 8$. The search engine weighting value for the search engine 24, after rounding, is $5*0.73 + 5*0.27 = 5$. The search engine weighting value for the search engine 26, after rounding, is $2*0.73 + 9*0.27 = 4$.

[1037] It will be apparent to those skilled in the art that the disclosed embodiments may be modified in numerous ways and may assume many embodiments other than the particular forms specifically set out and described herein. For example, the acts depicted in FIGS. 2 and 4 are not limited to the order depicted therein, and may be performed either in an alternative order or in parallel.

[1038] Accordingly, the above disclosed subject matter is to be considered illustrative, and not restrictive, and the appended claims are intended to cover all such modifications, enhancements, and other embodiments which fall within the true spirit and scope of the present invention. Thus, to the maximum extent allowed by law, the scope of the present invention is to be determined by the broadest permissible interpretation of the following claims and their equivalents, and shall not be restricted or limited by the foregoing detailed description.